

Anwar Shaikh

+1(404) 948-8990 | Atlanta, Georgia | anwarshaikh078@gmail.com | [Website](#) | [LinkedIn](#) | [Medium](#) | [Github](#)

EDUCATION

Georgia State University, Atlanta

August 2022 – July 2023

Master of Information Systems, Big Data Analytics

Dr. Babasaheb Ambedkar Marathwada University

June 2015 – May 2019

Bachelor of Engineering, Computer Science

TECHNICAL SKILLS AND CERTIFICATIONS

Programming Language: Python, SQL.

Google Cloud Services: GCS, Cloud Functions, PubSub, Composer, Dataflow, Dataproc, Bigquery, CloudSQL, Cloud Spanner, Compute Engine, VPC.

AWS: RDS, Redshift, Lambda, S3, EC2

Databases: Teradata, MySQL.

BigData: Hadoop, MapReduce, Spark, Kafka, Airflow, BeamSDK.

Others: Git, Gitlab, Jira

Certifications: Google Certified Professional Data Engineer, Apache Spark 3 - Real time Stream Processing using Python

PROFESSIONAL EXPERIENCE

Datametica Solutions Pvt Ltd, Pune, India

Verizon, Data Engineer 2

November 2021 – August 2022

- Developed a Pyspark(Spark) script to load data from Cloud Storage to Bigquery. The configuration was done on Dataproc, and achieved 99% job accuracy.
- Implemented CICD process to succeed automated deployments to the production environment. The entire project space had 23 tracks, we acquire all of them under the CICD umbrella. The manual deployments stopped, and prod deployments happened more frequently.
- Managed a group of 3 people, while working on CICD. Co-ordinated with them to resolve issues and communicate the needs to the client.

ADT Security Services, Data Engineer 1

July 2020 – November 2021

- Build Dataflow pipelines in python to fetch data from Relational Database to Bigquery. In the process also, performed data cleaning, validation and transformation.
- The user's Batch Load was delayed every day by 50–70 min, eventually delaying the SLA time. Suggested increasing the slots leveraging Reservation API. Deployed Cloud function to accomplish this and the orchestration was done in cloud composer. Reduced the overall SLA time by 45 minutes.
- Designed and developed a DAG generation framework in python to create 700+ DAGs. Helped save manual work for 15 days.
- Improved the performance of Airflow DAG's. One Dag used to take 13 hours, reduced it down to 3.5 hours.
- Incorporated stored procedure to attain SCD-1 and SCD-2. The same stored procedures consumed 3 hours in traditional database, optimized it to 55 minutes.

GameStop, Associate Data Engineer

August 2019 – July 2020

- Engineered the streaming data pipeline utilizing google cloud components. Processed 0.3 million json messages and loaded them to the stream tables. The data was persisted in the storage.
 - Programmed an Audit Framework to implement change data capture (CDC) process. This helped the team to keep track of the daily data capture, jobs status and later publish the report to the higher management.
 - Worked on Data Extraction from Informatica files, helped team save manual efforts to check each informatica workflow and mappings.
-

AWARDS AND ACHIEVEMENTS

- Hulk award, Rising Star award during the 3 years tenure at Datametica.
 - Principal Representative for the year 2018 – 2019.
 - Head of Department Representative Computer Science Department for 2017-2018.
-

VOLUNTEERING AND LEADERSHIP

- Student Co-Ordinator at annual cultural fest. Organized many events in across multiple fests during graduation.
 - Volunteered to work with training and placement cell in second and third year. Helped companies find the right candidates.
-